Week 4 Lab - Sentiment Analysis II

Clarissa Boyajian

2022-04-24

```
raw_tweets <- read.csv("https://raw.githubusercontent.com/MaRo406/EDS_231-text-sentiment/main/dat/IPCC_
                       header = TRUE)

# load sentiment lexicons
bing_sent <- get_sentiments('bing')
nrc_sent <- get_sentiments('nrc')
```

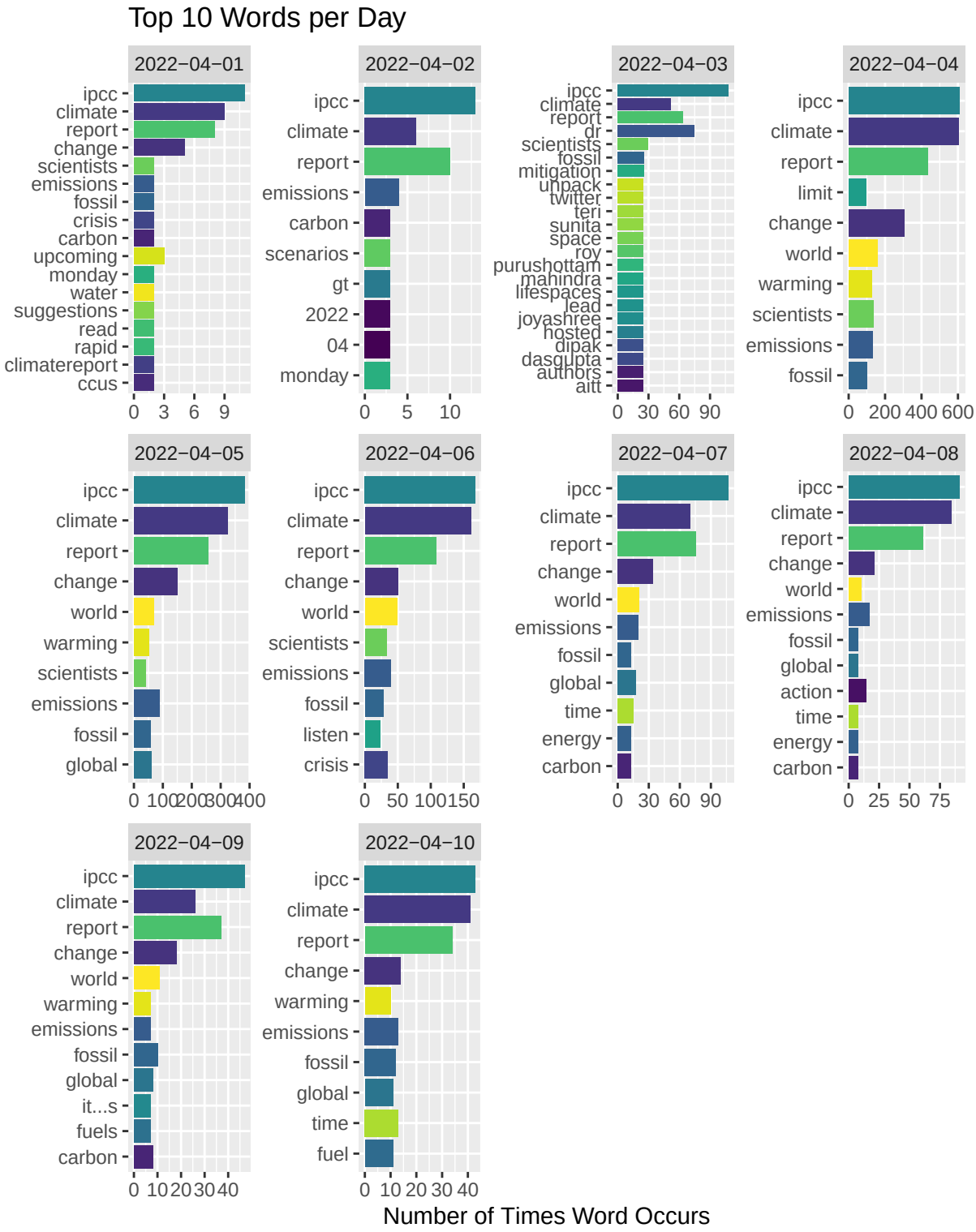
Question 1: Clean Data

```
clean_data <-
  raw_tweets[, c(4, 6, 10:11)]

tweets <-
  tibble(id = seq(1:length(clean_data$Title)),
         text = clean_data$Title,
         date = as.Date(clean_data$Date, '%m/%d/%y'),
         sentiment = clean_data$Sentiment,
         emotion = clean_data$Emotion) %>%
  mutate(text = str_replace(string = text,
                            pattern = "http.*[:space:]",
                            replacement = ""),
         text = str_replace(string = text,
                            pattern = "http.*$",
                            replacement = ""),
         text = str_replace(string = text,
                            pattern = "@.*[:space:]",
                            replacement = ""),
         text = str_replace(string = text,
                            pattern = "@.*$",
                            replacement = ""),
         text = str_to_lower(text))

words <- tweets %>%
  select(id, date, text) %>%
  unnest_tokens(output = word,
                input = text,
                token = "words") %>%
  anti_join(stop_words, by = "word") %>%
  left_join(bing_sent, by = "word") %>%
  left_join(
```

```
    tribble(
      ~sentiment, ~sent_score,
      "positive", 1,
      "negative", -1),
    by = "sentiment")
```

Question 2: Compare 10 Most Common Words per Day

```
words_common <- words %>%
  group_by(date, word) %>%
  summarise(count = n()) %>%
  group_by(date) %>%
  slice_max(count, n = 10)

ggplot(data = words_common,
       aes(x = count,
           y = reorder(word, count))) +
  geom_col(aes(fill = word)) +
  facet_wrap(~date, scales = "free") +
  guides(fill = "none") +
  scale_fill_viridis_d() +
  labs(x = "Number of Times Word Occurs",
       y = "",
       title = "Top 10 Words per Day")
```


Top 10 Words per Day
2022−04−01
ipcc
climate
report
change
scientists
emissions
fossil
crisis
carbon
upcoming
monday
water
suggestions
read
rapid
climatereport
ccus
0 3 6 9
2022−04−02
ipcc
climate
report
emissions
carbon
scenarios
gt
2022
04
monday
0 5 10
2022−04−03
ipcc
climate
report
dr
scientists
fossil
mitigation
unpack
twitter
teri
sunita
space
roy
purushottam
mahindra
lifespaces
lead
joyashree
hosted
dipak
dasgupta
authors
aitt
0 30 60 90
2022−04−04
ipcc
climate
report
limit
change
world
warming
scientists
emissions
fossil
0 200 400 600
2022−04−05
ipcc
climate
report
change
world
warming
scientists
emissions
fossil
global
0 100 200 300 400
2022−04−06
ipcc
climate
report
change
world
scientists
emissions
fossil
listen
crisis
0 50 100 150
2022−04−07
ipcc
climate
report
change
world
emissions
fossil
global
time
energy
carbon
0 30 60 90
2022−04−08
ipcc
climate
report
change
world
emissions
fossil
global
action
time
energy
carbon
0 25 50 75
2022−04−09
ipcc
climate
report
change
world
warming
emissions
fossil
global
it...s
fuels
carbon
0 10 20 30 40
2022−04−10
ipcc
climate
report
change
warming
emissions
fossil
global
time
fuel
0 10 20 30 40
Number of Times Word Occurs

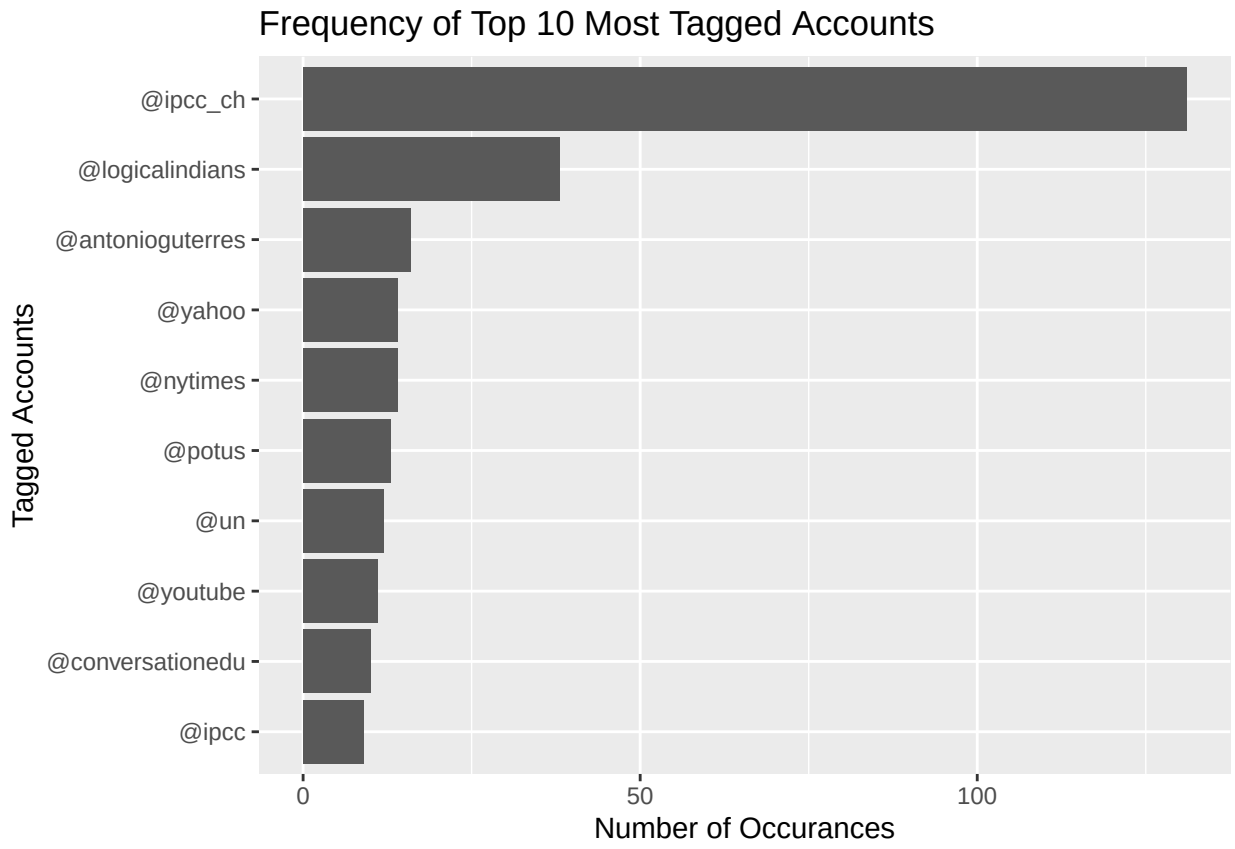
Question 3: Add Color to Wordcloud

```
words %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("darkgreen", "red", "lightgrey"),
                   max.words = 100)
```

Question 4: Most Tagged Accounts

```
corpus <- corpus(clean_data$Title)

tagged_tweets <- tokens(corpus, remove_punct = TRUE) %>%
  tokens_keep(pattern = "@*") %>%
  dfm() %>%
  textstat_frequency(n = 10)

tagged_tweets_tidy <- tokens(corpus, remove_punct = TRUE) %>%
  tokens_keep(pattern = "@*") %>%
  dfm() %>%
  tidy()
```

```
ggplot(data = tagged_tweets,
       aes(x = frequency,
           y = reorder(feature, frequency))) +
  geom_col() +
  labs(x = "Number of Occurances",
       y = "Tagged Accounts",
       title = "Frequency of Top 10 Most Tagged Accounts")
```



Questeion 5